

TEST SCORE ANALYSIS VIA THE STATISTICAL PROGRAM PSPP

Kirina Boykova, Denitza Charkova, Vanya Ivanova

Abstract: The following article examines the importance of analyzing test results to check students’ progress and suggest strategies for improvement. It also provides a detailed step-by-step tutorial on how to use the free statistical software PSPP, as well as how to read and interpret test results.

Key words: test, scores, statistical software, PSPP

1. Introduction

Inspired by Shotlekov’s article *From pilot to final online test version in Information technology* [5] this article gives a step-by-step guide on how to use data analysis software to assess a test in English for Specific Purposes class. It also tries to elaborate on the indicators the program uses and how to read the actual results.

It is often that educators use test scores solely to determine a student’s midterm or final grade irrespective of his/her overall achievements, knowledge or skills. A test, however, is able to provide a lot more information than only students’ results and achievements. It assesses the quality of the test itself. More precisely, such statistical analysis can reveal a lot about how individual students scored on each question, whether questions are reliable and which of them need to be revised or removed.

As St. John and Johnson argue, “the advantages of using qualitative data analysis software include being freed from manual and clerical tasks, saving time, being able to deal with large amounts of qualitative data, having increased flexibility, and having improved validity and auditability of qualitative research” [9]

Contemporary educators utilize numerous statistical programs to save time and efforts, and to provide credible data on the test quality. In this paper, we describe the features of PSPP, a statistical software provided by Gnu Software [10], which is the free alternative of the popular and massively used SPSS. [11] The text gives basic guidelines on how to use the program and explains some statistical analysis components that can easily be adopted even by newbie teachers.

2. Case study

For the purposes of this article, we have used the midterm test results of three groups of part-time students enrolled in the following majors: Business Information Technology, Informatics, and Software Technology and Design. All the students in the groups have scores attesting for level A1 – A2 [6]. The analysis is performed on the results of 57 people, 19 in each group after the first part-time trimester of academic year 2016/2017.

The midterm test included 21 questions with a maximum result of 38 points. We chose to include three types of questions: multiple choice items (thirteen of them in the test), true/false items (two of them in the test), matching items (six of them in the test). The students had 30 minutes to complete the test.

For the assessment we used the Google add-on *Flubaroo* which is a free tool that helps you quickly grade multiple-choice or fill-in-blank assignments. [3] It works by comparing the students' answers to the answer key provided by the tutor and is, in fact, incredibly convenient when working with multiple choice items.

3. Terminology

Prior to the PSPP guide we would like to clarify some of the terms which later appear as we discuss the results' values. The following list is not by any chance complete or thorough; we only wanted to focus on the basics so anyone could use the tool as the mission of the article is to make statistical tools more accessible for tutors with no prior experience or expertise.

- Bar chart – a figure made up of columns plotted on a graph, where the columns are positioned over a label that represents a categorical variable [7];
- Histogram – a figure showing the characteristics of the distribution of a survey; each measure is represented by a single block that is placed over the midpoint of the class interval into which the measure falls [1];

- Kurtosis – a measure of whether the data is peaked or flat relative to a normal distribution [5];
- Percentile – a data point below which a given percentage of observations in a group falls [4];
- Pie chart – a type of figure that displays data in a circular graph. The pieces of the graph are proportional to the fraction of the whole in each category;
- Range – the distance between the highest and lowest score; numerically, the range equals the highest score minus the lowest score [1];
- Mean – the average result of a test, survey, or experiment;
- Median – the score that divides the results in half – the middle value;
- Mode – the most common result (the most frequent value), the score that occurs most frequently;
- Skewness – a measure of symmetry, or the lack of symmetry [2];
- Standard deviation – a measure of the spread of scores within a set of data [7];
- Variance – it gives a measure of how the data distributes itself about the mean or expected value; the variance looks at all the data points and then determines their distribution [8];
- Z-score – indicates how far and in what direction an item deviates from its distribution’s mean [1].

4. PSPP Tutorial

Before using the free statistical software, the user needs to download the program (for the preferred operating system) and prepare the data. The imperative data for the analysis is student’s identification and test score. It is also important to remark that the format of the data file should be .txt or .csv. Once the user completes these preconditions, it is time for the actual procedure. Step 1 is data import, which is done from the *File* menu. Step 2 is selecting the *Variable* view option which will assist the user in the data preparation. Step three is to making sure that the program reads the scores properly. For this PSPP needs *Numeric* variable type and *Nominal* variable measure (see Figure 1).

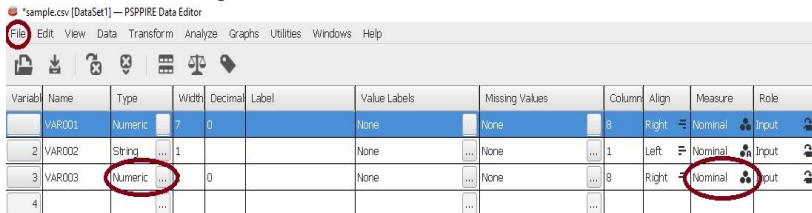


Figure 1.

Since the program has various analysis options, we are only showing the Descriptive statistics selections which serve the needs of our test analysis. For the next step, select the Analyze menu, choose Descriptive statistics, Frequencies and then pick the suitable indicators. For the final step, mark the variable that you want to analyze (1.1), then choose the indicators from the list (1.2); choose the figure that you want to observe with the results – a histogram (1.3) as shown on Figure 2.

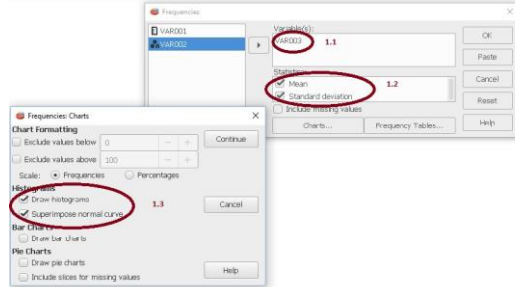


Figure 2.

All results of the analysis will appear in a new window along with figures and charts. In case a Warning message appears, once you start with the analysis, you need to return to the *Variable* view window to eliminate a potential discrepancy. Usually, it has to do with the *Type* and *Measure* of the variables.

5. Results

Since we do not have data from a preliminary test to compare to the midterm test, the analysis will not be able to give us information on the prospective progress of the students or if the test has been bettered. In this case, we can only analyze scores between the groups and the scores of the sample. In this study, the three groups cover the same educational content and the analysis is based on classes with same duration and number of sessions (the students who study Informatics had an extra class which was dedicated to additional activities unrelated to the core material such as discussions and listening exercises). At the end of the trimester the three groups had an online based final test, the result of which contributed to their midterm grade with 30% (other 30% came from attendance and 40 % came from individual assignments and homework). [6]

We used the raw scores of the test in PSPP and collected the following results for further analysis: we first got the descriptive statistics for the sample, such as the *Frequency* (how often we have each of the results), the *Percent* (what percent of the students scored with this result) and the *Cumulative Percent* (what percent of the

students got the same or lower result; where the highest score encompasses 100% of all scores). These are shown on Fig. 3. We can also observe the histogram which showed the distribution of the results on Fig. 4.

VAR003					
Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	10	1	1.75	1.75	1.75
	12	1	1.75	1.75	3.51
	13	1	1.75	1.75	5.26
	15	1	1.75	1.75	7.02
	16	1	1.75	1.75	8.77
	17	3	5.26	5.26	14.04
	18	2	3.51	3.51	17.54
	19	1	1.75	1.75	19.30
	20	3	5.26	5.26	24.56
	22	2	3.51	3.51	28.07
	23	4	7.02	7.02	35.09
	24	5	8.77	8.77	43.86
	25	5	8.77	8.77	52.63
	26	3	5.26	5.26	57.89
	27	1	1.75	1.75	59.65
	29	4	7.02	7.02	66.67
	30	5	8.77	8.77	75.44
	31	3	5.26	5.26	80.70
	32	5	8.77	8.77	89.47
	33	1	1.75	1.75	91.23
	35	2	3.51	3.51	94.74
	36	1	1.75	1.75	96.49
	37	2	3.51	3.51	100.00
<i>Total</i>		57	100.0	100.0	

Figure 3.

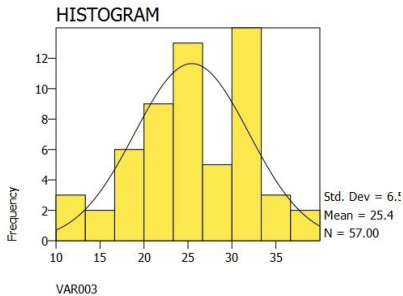


Figure 4.

We should also observe the measure of symmetry (*Skewness*) – it could be symmetrical and centered or with a long tail to the right (positively skewed/asymmetrical), or with a long tail to the left (negatively skewed/asymmetrical). Other valuable information could be withdrawn from the average result of the sample (*Mean*), the *Standard deviation* (the smaller the number is, the more homogenous the results), and the *Kurtosis* which quantifies whether the shape of the data distribution matches the normal distribution – a peaked curve signifies positive kurtosis and a flatten one means the kurtosis is negative. [4] The *Range* is also a feature of the descriptive statistics although it could not give us particularly valuable

information as it is the value of the highest minus the lowest results each of which could be an anomaly (see Figure 5).

VAR003

<i>N</i>	<i>Valid</i>	57
	<i>Missing</i>	0
<i>Mean</i>		25,44
<i>Mode</i>		.
<i>Std Dev</i>		6,51
<i>Kurtosis</i>		-,48
<i>S.E. Kurt</i>		,62
<i>Skewness</i>		-,30
<i>S.E. Skew</i>		,32
<i>Range</i>		27,00
<i>Minimum</i>		10,00
<i>Maximum</i>		37,00
	50 (Median)	25

Figure 5.

The heretofore analysis showed us data about the sample – how all the students have performed, what are the average, minimum and maximum scores, as well as how most scores are distributed. The software, however, could be truly practical in case we want to compare the different groups. Figure 6 shows us what is the *Standard deviation* in each of the groups, what are the score ranges within them and how they refer to the sample.

Descriptives

		<i>N</i>	<i>Mean</i>	<i>Std. Deviation</i>	<i>Std. Error</i>	<i>95% Confidence Interval for Mean</i>		<i>Minimum</i>	<i>Maximum</i>
						<i>Lower Bound</i>	<i>Upper Bound</i>		
<i>VAR003</i>	<i>B</i>	19	26,16	5,37	1,23	23,57	28,74	13	36
	<i>I</i>	19	24,37	6,14	1,41	21,41	27,33	15	35
	<i>S</i>	19	25,79	7,96	1,83	21,96	29,62	10	37
	<i>Total</i>	57	25,44	6,51	,86	23,71	27,16	10	37

Figure 6.

6. Conclusions

As we observe the results we should mind out that the mean as a representative of the central tendency measures is sensitive to radical scores. In our case the mean of the sample is approximate to the values we have for the different groups; thus, we can consider it a reliable indicator of the central tendency. We should also state that the normal distribution is partially depending on the size of the sample – the less the people in the sample the more ambiguous the distribution. The normal distribution is also predetermined by the spread out of the figure and its peaked or flattened shape –

observing the *Skewness* and *Kurtosis* we can declare that the histogram is negatively skewed – we have a higher number of higher scores on the test. As we study the *Standard deviation* of the sample we can state that most results are centered 6,51 units around the *Mean*.

A more accurate idea of the sample performance can be acquired by observing the three groups' scores and how they refer to the sample score. Comparing the *Standard deviation* of the three separate groups, we can state that they had similar levels of variability. We can first consider the *Mean* of the three groups – the highest deviation from the sample has the second group (Informatics), where we can also notice the slightest difference between the minimum and maximum score and a *Standard deviation* closest to that of the sample. The group with the least *Standard deviation*, thus with the most homogenous group results appears to be the Business information technology majors. Software technology and design display the highest *Standard deviation*, accordingly the highest difference between a minimum and maximum score with a *Mean* closest to that of the sample. As a conclusion, we can state that the sample demonstrates a rather heterogenous language levels at the midterm test. The Software technology and design group accounted for the highest and the lowest score in the sample and the sample itself scored closer to the higher grade.

We will consider the sample survey for further research once the second semester is completed and we are able to compare these scores to the final ones. As a notice for future research we will consider testing a larger sample (if possible).

Acknowledgements

This paper is partially supported by the IT15-FMIIT-004 project of the Scientific Fund of the Plovdiv University “Paisii Hilendarski”, Bulgaria.

Literature

- [1] <https://www.animatedsoftware.com/statglos/statglos.htm#index> (last accessed: 25.10.2016)
- [2] <http://atomic.phys.uni-sofia.bg/local/nist-e-handbook/e-handbook/eda/section3/eda35b.htm> (last accessed: 25.10.2016)
- [3] Flubaroo User Guide, <http://www.flubaroo.com/flubaroo-user-guide> (last accessed: 25.10.2016)
- [4] Kalinov, K., Statistical methods in Behavioral and Social Sciences, New Bulgarian University, Sofia, 2013, p. 26–27.

- [5] Shotlekov, I., From pilot to final online test version in Information technology, <http://diuu.bg/emag/2919/3/> (last accessed: 25.10.2016)
- [6] Shotlekov, I., et al, Boosting efficiency of project-oriented teaching and learning through classroom management and online testing, *Mathematics and Informatics*, Vol. 56, No. 5, 2013 p.478 – 488
- [7] <https://statistics.laerd.com/statistical-guides/measures-of-spread-standard-deviation.php> (last accessed: 25.10.2016)
- [8] <http://stattrek.com/statistics/charts/histogram.aspx?Tutorial=AP> (last accessed: 25.10.2016)
- [9] St John W., Johnson P., The pros and cons of data analysis software for qualitative research, *J Nurs Scholarsh*, 2000, 32 (4), p. 393 – 397
- [10] <https://www.gnu.org/software/pspp/> (last accessed: 25.10.2016)
- [11] <http://www.ibm.com/analytics/us/en/technology/spss/> (last accessed: 25.10.2016)

Факултет по математика и информатика

Пловдивски университет

Бул. „България“ 236,

4003 Пловдив, България

e-mail: kirina.boykova@gmail.com, d.charkova@gmail.com, vantod@abv.bg

АНАЛИЗ НА ТЕСТОВИ РЕЗУЛТАТИ ЧРЕЗ СТАТИСТИЧЕСКАТА ПРОГРАМА PSPP

Кирина Бойкова, Деница Шаркова, Ваня Иванова

Резюме: Настоящата статия изследва анализа на тестови резултати на обучаеми с помощта на статистически софтуер с цел проверка на техния напредък, както и събиране на данни за последващи възможности за оптимизиране на качеството на обучение. Статията представя и подробно описание за използване на програмата за съответните цели, както и насоки как да бъдат разчетени и интерпретирани получените данни.

Ключови думи: тест, резултати, статистически софтуер, PSPP